AD NO. ____________________________
DTC PROJECT NO. 8-CO-160-UXO-016
REPORT NO. ATC-9326

SHALLOW WATER

UXO TECHNOLOGY DEMONSTRATION SITE

SCORING RECORD NO. 4

SITE LOCATION:
U.S. ARMY ABERDEEN PROVING GROUND

DEMONSTRATOR:
CONCURRENT TECHNOLOGIES CORPORATION
100 CTC DRIVE,
JOHNSTOWN, PA 15904-1935

TECHNOLOGY TYPE/PLATFORM:
FEREX DLG-GPS MAGNETOMETER SYSTEM

PREPARED BY:
U.S. ARMY ABERDEEN TEST CENTER
ABERDEEN PROVING GROUND, MD 21005-5059

JANUARY 2007

Prepared for:
U.S. ARMY ENVIRONMENTAL COMMAND
ABERDEEN PROVING GROUND, MD 21010-5401

U.S. ARMY DEVELOPMENTAL TEST COMMAND
ABERDEEN PROVING GROUND, MD 21005-5055

DISTRIBUTION UNLIMITED, JANUARY 2007.

DISPOSITION INSTRUCTIONS

CSTE-DTC-AT-SL-E

MEMORANDUM FOR RECORD

SUBJECT: Operations Security (OPSEC) Review of Paper/Presentation

1. The attached record entitled "The Shallow Water UXO Technology Demonstration Site Scoring Record #4" dated January 2007 is provided for review for public disclosure in accordance with AR 530-1 as supplemented. The scoring record is proposed for public release via the internet.

2. I, the undersigned, am aware of the intelligence interest in open source publications and in the subject matter of the information I have reviewed for intelligence purposes. I certify that I have sufficient technical expertise in the subject matter of this report and that, to the best of my knowledge, the net benefit of this public release outweighs the potential damage to the essential secrecy of all related ATC, DTC, ATEC, Army or other DOD programs of which I am aware.

NAME (Printed)	SIGNATURE	DATE
Gary Rowe		January 2007

CONCURRENCE:	NAME (Printed)	SIGNATURE	DATE
Program Mgr/Customer (If not ATC owned technology)	Patrick McDonnell		30 JAN 07
Core Director	Charles Valz ~~William Burch~~		1/24/07 03 JAN 07
Core OPSEC Tech POC	Michael Karwatka		5 Jan 07
ATC OPSEC Officer/ Security Manager	William Murr		8 Feb 07
Public Affairs Spec	Susan Hagan		20 Feb 07
Tech Dir, ATC	John R. Wallace		24 Feb 07

(Return to ATC PAO for further processing)

DTC GO/SES	N/A		

Encl
as

ACKNOWLEDGMENTS

Author:

Gary W. Rowe
Military Environmental Technology Demonstration Center (METDC)
U.S. Army Aberdeen Test Center (ATC)
U.S. Army Aberdeen Proving Ground (APG)

Contributors:

William Burch
Military Environmental Technology Demonstration Center
U.S. Army Aberdeen Test Center
U.S. Army Aberdeen Proving Ground

Christina McClung
Aberdeen Data Services Team (ADST)
Logistics Engineering and Information Technology Company (Log.Sec/Tri-S)
U.S. Army Aberdeen Proving Ground

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
January 2007	Final	20 through 24 March 2006

4. TITLE AND SUBTITLE SHALLOW WATER UXO TECHNOLOGY DEMONSTRATION SITE SCORING RECORD NO. 4 (CTC, FEREX DLG-GPS, MAG)	**5a. CONTRACT NUMBER** **5b. GRANT NUMBER** **5c. PROGRAM ELEMENT NUMBER**
6. AUTHOR(S) Rowe, Gary	**5d. PROJECT NUMBER** 8-CO-160-UXO-016 **5e. TASK NUMBER** **5f. WORK UNIT NUMBER**
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Commander U.S. Army Aberdeen Test Center ATTN: CSTE-DTC-AT-SL-E Aberdeen Proving Ground, MD 21005-5059	**8. PERFORMING ORGANIZATION REPORT NUMBER** ATC-9326
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Commander U.S. Army Environmental Command ATTN: SFIM-AEC-ATT Aberdeen Proving Ground, MD 21005-5401	**10. SPONSOR/MONITOR'S ACRONYM(S)** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)** Same as Item 8

12. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution unlimited.

13. SUPPLEMENTARY NOTES

None.

14. ABSTRACT

This report documents the efforts of Concurrent Technologies Corporation (CTC) to detect and discriminate inert unexploded ordnance (UXO) using a FEREX DLG-Global Positioning System (GPS) magnetometer (MAG). Testing was conducted at ATC Standardized Shallow Water UXO Technology Demonstration Site. A description of the tested system and an estimate of survey costs along with the analysis of the system performance are provided.

15. SUBJECT TERMS

CTC, UXO Standardized Technology Demonstration Site, Shallow Water, FEREX DLG-GPS MAG, MEC

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
Unclassified	Unclassified	Unclassified	SAR		

TABLE OF CONTENTS

SECTION 4. TECHNICAL PERFORMANCE RESULTS

SECTION 5. APPENDIXES

SECTION 1. GENERAL INFORMATION

1.1 BACKGROUND

Technologies under development for the detection and discrimination of munitions and explosives of concern (MEC), i.e., unexploded ordnance (UXO) and discarded military munitions (DMM), require testing so their performance can be characterized. To that end, the U.S. Army Aberdeen Test Center (ATC) located at Aberdeen Proving Ground (APG), Maryland, has developed a Standardized Shallow Water Test Site. This site provides a controlled environment containing varying water depths, multiple types of ordnance and clutter items, as well as navigational and detection challenges. Testing at this site is independently administered and analyzed by the government for the purposes of characterizing technologies, tracking performance during system development, and comparing the performance and costs of different systems.

The Standardized UXO Technology Demonstration Site Program is a multiagency program spearheaded by the U.S. Army Environmental Command (USAEC). ATC and the U.S. Army Corps of Engineers Engineering, Research and Development Center (ERDC) provide programmatic support. The Environmental Security Technology Certification Program (ESTCP), the Strategic Environmental Research and Development Program (SERDP), and the Army Environmental Quality Technology Program (EQT) provided funding and support for this program.

1.2 OBJECTIVE

The objective of the Shallow Water Standardized UXO Technology Demonstration Site is to evaluate the detection and discrimination capabilities of existing and emerging technologies and systems in a shallow water environment. Specifically:

a. To determine the demonstrator’s ability to survey a shallow water area, analyze the survey data, and provide a prioritized “Target List” with associated confidence levels in a timely manner.

b. To determine both the detection and discrimination effectiveness under realistic scenarios that varies ordnance, clutter, and bathymetric conditions.

c. To determine cost, time, and manpower requirements needed to operate the technology.

1.3 CRITERIA

The scoring criteria specified in the Environmental Quality Technology - Operational Requirements Document (EQT-ORD) (app D, ref 1) for: A(1.6.a): UXO Screening, Detection and Discrimination document are presented in Table 1-1. Very little information was available on the capabilities of shallow water detection systems when these criteria were developed. However, they were used in the design of the test site, and the five metrics were used to measure system performance in this report.

TABLE 1-1. SCORING CRITERIA

Metric	Threshold	Objective
Detection	80% ordnance items buried to 1 foot and under 8 feet (2.4 m) of water at a standardized site detected	95% ordnance items buried to 4 feet and under 8 feet (2.4 m) of water at a standardized site detected
Discrimination	Rejection rate of 50% of emplaced non-UXO clutter at a standardized site with a maximum false negative rate of 10%	Rejection rate of 90% of emplaced non-UXO clutter at a standardized site with a maximum false negative rate of 0.5%
Reacquisition	Reacquire within 1 meter	Reacquire within 0.5 meter
Cost rate	$4000 per acre	$2000 per acre
Production rate	5 acres per day	50 acres per day

The ATC shallow water site was designed to evaluate the threshold detection level of a range of ordnance at the 1-foot + 8-foot requirement. Limited information is available at the objective detection level. All other measured results will be evaluated against both criteria levels.

1.4 APG SHALLOW WATER SITE INFORMATION

1.4.1 Location

The Aberdeen Area of APG is located in the northeast portion of Maryland on the western shore of the Chesapeake Bay in Harford County. The Shallow Water Test Site is located within a controlled range area of APG.

1.4.2 Soil Type

The area chosen for the shallow water test site was known as Cell No. 3 in a dredge-spoil field. The cell bottom is primarily composed of sediment removed from the Bush River. This is a freshwater site.

1.4.3 Test Areas

a. The test site contains five areas: calibration grid, blind test grid, littoral, open water, and deeper water. Additional detail on each area is presented in Table 1-2. A schematic of the calibration lanes is shown in Figure 1.

TABLE 1-2. TEST AREAS

Area	Description
Calibration grid	The calibration area contains 15 projectiles, 3 each 40, 60, 81, 105, and 155 mm. One of each projectile type is buried at the projectile diameter to depth ratio shown in Figure 1. This area is designed to provide the user with a sensor library of detection responses for the emplaced targets and an understanding of their resistivity prior to entering the blind test fields. Two "clutter-cloud" target scenarios have been constructed adjacent to this area (fig. 1).
Blind grid	The blind grid contains 644 detection opportunities. Each grid cell is 2 x 2 m^2. At the center of each cell is either an ordnance item, clutter, or nothing. Surrounding the blind grid on three sides are 3.6-kg (8-lb) shot puts, buried 0.3 meter deep in the sediment. The shot puts can be used as a navigational/Global Positioning System (GPS) check. The GPS coordinates for the center of each grid and the shot put locations are provided to the vendor prior to testing.
Littoral	This is a sloping area on one side of the pond with vegetation growing into the water line. Water depth ranges from 0.3 to 1.8 meters. It contains a variety of navigational and detection challenges.
Open water	The open water scenario contains a variety of navigational, detection, and discrimination challenges. Water depth varies from 1.8 to 3.4 meters.
Deeper water	The water depth in this area varies between 3.4 and 4.3 meters.

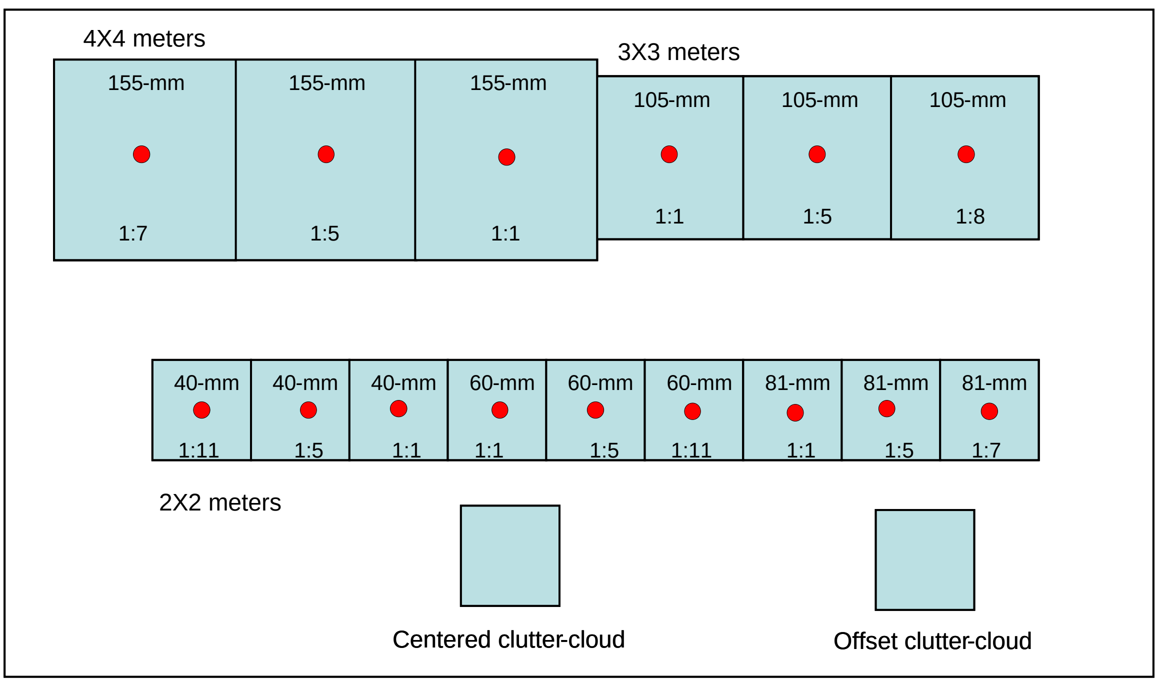

Figure 1. Schematic of the calibration grid.

b. The water depth at this facility during testing is maintained such that the calibration and blind grid areas meet the 2.4-meter (8-ft) detection criterion specified in paragraph 1.3. The test site is approximately 2.8 hectares (6.9 acres) in size.

1.5 GROUND TRUTH TARGETS

The ground truth is comprised of both inert ordnance and clutter items. The inert ordnance items are listed in Table 1-3. All items were located in storage sites at APG. The items have not been fired or degaussed.

Clutter items fit into one of three categories: ferrous, nonferrous, and mixed-metals. The ferrous and nonferrous items have been further divided into three weight zones as shown in Table 1-4 and distributed throughout all test areas. Most of this clutter is comprised of ordnance components; however, there are also industrial scrap metal and cultural items as well. The mixed-metals clutter is comprised of scrap ordnance items or fragments that have both a ferrous and nonferrous component and could reasonably be encountered in a range area. The mixed-metals clutter was placed in the open water area only.

TABLE 1-3. INERT ORDNANCE TARGETS

Description	Length, mm	Diameter, mm	Aspect Ratio, W/L	Weight, g
40-mm L70 projectile	208	40	0.1923	965
60-mm mortar M49A2	185	60	0.3243	975
81-mm mortar M374	528	81	0.1534	3969
81-mm mortar M821	510	81	0.1588	3338
105-mm projectile M1	445	105	0.2360	13834
155-mm M107 projectile	684	155	0.2266	41731
8-in. M104/106	856	203	0.2371	89811

TABLE 1-4. CLUTTER WEIGHT RANGES

Clutter Type	Weight Range in Grams		
	Small	Medium	Large
Ferrous	10 to 510	511 to 2200	> 2201
Nonferrous	10 to 270	275 to 800	> 801

SECTION 2. SYSTEM UNDER TEST

2.1 DEMONSTRATOR INFORMATION

Concurrent Technologies Corporation (CTC), as part of their Broad Agency Announcement (BAA) submittal (app D, ref 2), provided the information in sections 2.2 through 2.7 in their technical management plan. ATC's comments on the demonstrated system are provided in section 2.8.

2.2 SYSTEM DESCRIPTION

The Foerster system that CTC used at the shallow water test site is a commercial off-the-shelf system that has been used in shallow waters successfully at numerous jobs in North America, Europe, and Asia. The system that was demonstrated at the ATC as a proof of concept used four sensors. However, it is scalable to be larger and has most recently been used in Tokyo Bay to locate UXO using a 16-sensor array.

CTC proposes a fluxgate vertical gradient magnetic sensor technology coupled with differential global positioning methods, specifically, the Foerster FEREX® 4.032 geophysical sensor coupled with the Trimble 5700 Differential Global Positioning System (DGPS) technology. The proposed FEREX® device uses fluxgate vertical gradient magnetic technology to facilitate the detection and discrimination of ferrous metallic objects. Ferromagnetic parts that are located in the Earth's magnetic field generate a magnetic interference field in their environment. This interference field can be detected using the Foerster differential magnetometer. Its amplitude and its magnetic polarity are displayed and can be used for object pinpointing. The operator can choose from eight linear measurement range settings (from 0 to 3 up to 0 to 1000 nT) and one logarithmic measurement range setting on the instrument. The unit displays a 0.3-nT resolution and will use four separate detection probes. The FEREX 4.032 sensor can be used in the data logger versions together with the FEREX-DATALINE® software for computer-assisted cartography and localization.

FEREX-DATALINE® 4.800 software is the analysis software that runs under Microsoft Windows for interactive, graphical evaluation of measurements to calculate object coordinates and positioning as well as the size and depth of suspected ferromagnetic objects. DATALINE enables exact scaled reproduction of recorded and measured data by means of color-coded magnetic field value charts. ISO lines or three-dimensional presentations can be displayed to additionally optimize the presentation of measurements. Data exports are possible with a selectable delimiter as a file for further editing or evaluation in other application programs. CTC intended to use the FEREX DLG with GPS data logger in the four-sensor configuration for the shallow water demonstration where applicable. Operator controls and indicators are within the unit housing and within the operator's field of view; the battery pack is integrated in the carrying tube; and a permanently integrated loudspeaker within the detector assists with defining the survey parameters and warns the operator of unacceptable DGPS quality. Figure 2 shows the electronic schematic of what CTS proposed.

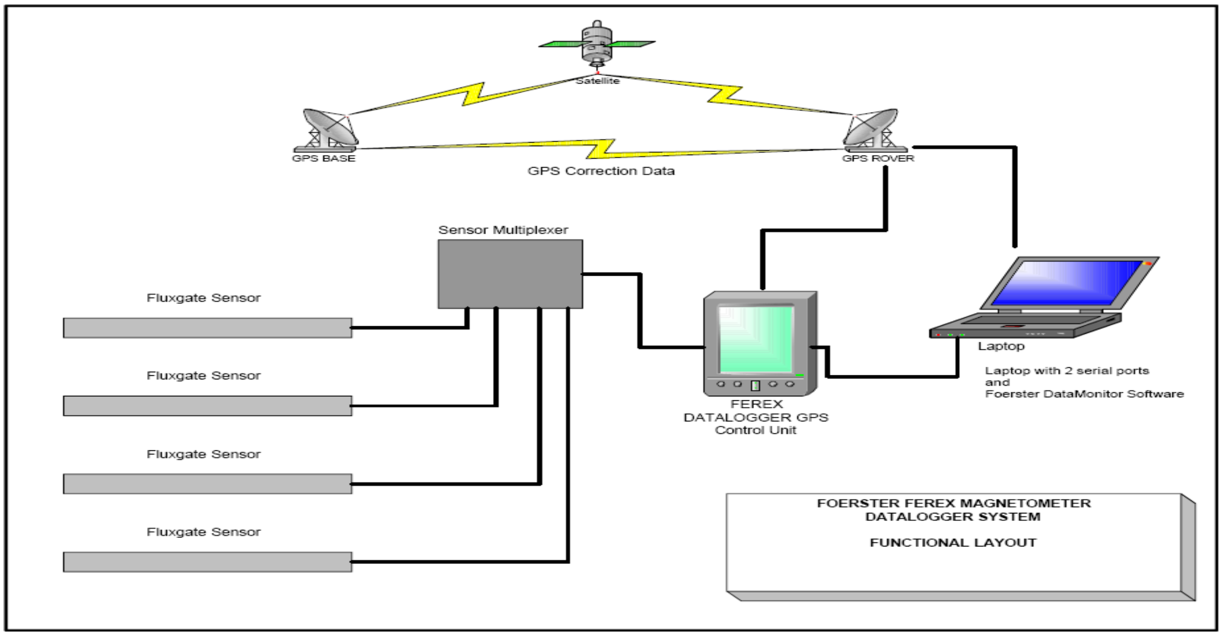

Figure 2. CTC system schematic.

2.3 DEMONSTRATOR'S POC AND ADDRESS

POC: Mr. Josh Bowers
email: bowersr@ctc.com

Address: Concurrent Technologies Corporation
100 CTC Drive
Johnstown, PA 15904-1935

2.4 DEMONSTRATOR'S SITE SURVEY METHOD

The shallow water demonstration area was approximately 6.9 acres in size and had depths ranging from 0.3 to 4.3 meters. These features were used to evaluate the Foerster geophysical system performance under these conditions. Because of the lack of tall, dense vegetation at the site, the DGPS was integrated with the FEREX 4.032 geophysical sensor as a boat-mounted system (fig. 3). For this demonstration, a transect sensor spacing of no more than 0.50 meter was required when using the proposed geophysical sensor to detect and discriminate objects as small as 40-mm projectiles. On the basis of the FEREX data logger's ability to guide the operator on straight acquisition lines and the development of rigorous field procedures for the field crew, it was expected that adequate transect spacing would be maintained under all conditions.

To collect the best possible data, CTC took depth soundings of the survey area to optimize depth settings for the sensors used. The proposed navigation and data collection procedures have been proved effective under the types of conditions anticipated at the shallow water demonstration area. It was CTC's goal to maximize the efficiency of the acquisition process while minimizing the potential for data gaps and missed targets of interest.

Figure 3. CTC shallow water UXO detection platform.

2.5 DEMONSTRATOR'S QC AND QA

a. Field personnel, data processors, and data interpreters implemented the QC program in a consistent fashion. In general, the QC program consisted of a series of preproject tests, and once the project had started, a test regimen was applied for each acquisition session. The test

regimen included functional checks to ensure that the position and geophysical sensor instrumentation was functioning properly before and after each data acquisition session, processing checks to ensure that the data collected were of sufficient quality and quantity to meet the project objectives, and interpretation checks to ensure that the processed data were representative of the site conditions. Preproject tests included functional checks to ensure that the position and geophysical sensor instrumentation was operating within its defined parameters. Specific preproject tests included the following:

(1) Five-minute static tests for each FEREX 4.032 system.

(2) Cable integrity tests for each FEREX 4.032 system.

(3) Manufacturer-suggested functional checks for the DGPS.

(4) DGPS quality checks from the FEREX data logger screen.

b. Specific functional checks during the data acquisition program included the following:

(1) Sensor jig metal check (ensure no metal on acquisition personnel).

(2) Static position system check (accuracy and repeatability of position).

(3) Static geophysical sensor check (repeatability of measurements and influence of ambient noise).

(4) Static geophysical sensor check with a test item (repeatability and comparability of measurements with metal present).

(5) Kinematic geophysical sensor check with a test item (repeatability and comparability of measurements with sensor in motion).

(6) Repeatability of overall data (resurvey of a portion of the survey area during each data acquisition session).

(7) CTC reoccupied the survey monuments with the DGPS to ensure comparability, accuracy, and repeatability of the positioning systems.

c. The QA procedures applied during the processing phase of the project were performed each day in the field to ensure the integrity of the data. Data that were not of sufficient quality and quantity to meet the project objectives were documented and re-collected.

d. Procedural checks during the processing of the data included the following:

(1) Evaluation of the static position and FEREX 4.032 data. FEREX 4.032 static noise above a predefined threshold was documented, and a root cause analysis was performed before collecting additional data.

(2) Evaluation of the kinematic geophysical sensor check. These data allowed the processor to qualitatively and quantitatively monitor the noise level and repeatability of the data over a “standard” item as well as ensure that the data were merged correctly (i.e., the data contained no time or position shift, also known as “lag”).

(3) Corner buoy locations for the survey grid were compared with known survey data and verified.

(4) Sample density along transects was verified through statistics.

(5) Unreasonable FEREX 4.032 measurement values were documented and compared with the site cultural features map. Foerster developed internal software to meet some of the needs during merging, processing, and interpretation of the data.

e. Quality assurance measures applied during interpretation of the data included the following:

(1) Depth and target volume information was calculated by a “dipole fit” algorithm, based on a method that has been proved and accepted worldwide as a qualified tool for applications such as these.

(2) The target evaluation was performed on the basis of magnetic polarities, selected by the user.

(3) A quality indication informed the user how well the dipole fit method could be performed with the user’s selected polarity configuration.

(4) Normally, several above-ground metal features (e.g., fence posts, monitoring wells, etc.) are selected from each acquisition session for reacquisition by field personnel to verify the accuracy of the interpreted position coordinates. Such items were located in the vicinity of the shallow water demonstration area.

(5) Comparison of the position and FEREX 4.032 data with the site features map (e.g., above-ground cultural features were documented; should be variance in the track path). Interpreted data characteristics were compared with the known responses acquired during the initial test program (e.g., calibration lane).

f. In addition, CTC performed quality assurance on the data using the Geosoft software suite.

2.6 DATA PROCESSING DESCRIPTION

DGPS position data were acquired and recorded within the FEREX data logger at a rate of 1 Hz. The Foerster FEREX® data were recorded at 20 Hz by the internal data logger. The

FEREX requires GGA and LLK National Marine Electronics Association (NMEA) strings for defining positions and pulses per second as a timing constant.

Foerster DATALINE software was used to convert the FEREX data to units of nanotesla. The positioning and FEREX signal data were merged within the data logger during acquisition. The DATALINE software has been proved and verified on various UXO removal projects across the world. It is the standard software tool in numerous military units.

The FEREX raw data were output via the DATALINE software as an American Standard Code for Information Interchange (ASCII) file that contained the relative X/Y, a selected local (e.g., UTM), and WGS84 coordinates and the corresponding FEREX signal intensity reading. FEREX data were interpolated between corresponding position segments that were spaced at intervals of 12 to 18 inches along the ground surface, at a normal acquisition speed of 3 ft/sec on land, and it was anticipated that the data acquisition speed may have been slightly less with the motor and boat used. Samples along each acquisition transect were produced at intervals of approximately 1 to 3 inches over water.

2.7 DEMONSTRATOR'S SITE PERSONNEL

Project Geophysicist: Mr. Josh Bowers

Data Acquisition Specialists: Mr. Thomas Himmler
Mr. Myles Capen

2.8 ATC'S SURVEY COMMENTS

This is the only boat-mounted system that has been tested with the ability to vary the depth of the sensors with the water depth (fig. 4 and 5). Keeping the magnetometers a uniform depth from the bottom should provide a more consistent signal response, leading to better detection and discrimination results.

Having a variable sensor depth also increases the maneuverability and capability of the system as the water levels change.

Figure 4. CTC shallow water UXO detection platform - deep deployment.

Figure 5. CTC shallow water UXO detection platform - shallow deployment.

SECTION 3. SURVEY COST ANALYSIS

3.1 DATES OF SURVEY

The FEREX DLG-GPS magnetometer system was tested from 20 through 24 March 2006.

3.2 SITE CONDITIONS

3.2.1 Atmospheric Conditions

An ATC weather station located adjacent to the test site recorded the average temperature and precipitation on an hourly basis for each day of operation. The temperatures listed in Table 3-1 represent the average temperature from 0700 through 1700. The hourly weather logs used to generate this summary are provided in Appendix A.

3.2.2 Water Conditions

Water conditions were monitored using a TIDALITE IV Portable Tide Gauge System©. Data recorded included: water depth and temperature, significant wave height based on the average 1/3 wave height seen over the test period using the Draper/Tucker analysis method, and the full-wave frequency calculated by full-wave mean crossing detection. The values displayed in Table 3-1 were averaged from 0700 through 1700. The water conditions during the CTC survey were lost because of a malfunction in the portable tide gauge system. The water depth was measured against an elevation marker attached to the pier.

TABLE 3-1. SITE CONDITION SUMMARY

Date, 06	Air Temperature, °C	Wind, km/h	Water Temperature, °C	Water Depth, m[a]	Significant Wave Height, m	Wave Frequency, Hz
20 Mar	12.9	4.7	Lost	-0.1	Lost	Lost
21 Mar	8.1	1.2	Lost	-0.1	Lost	Lost
22 Mar	22.4	4.1	Lost	0.2	Lost	Lost
23 Mar	13.5	6.7	Lost	-0.2	Lost	Lost
24 Mar	18.5	4.5	Lost	-0.2	Lost	Lost

[a]Variance between the required 2.4-meter test depth and actual test conditions.
Lost = instrumentation malfunction.

3.3 SURVEY ACTIVITIES

The information contained in this section provides an estimate of the time needed and costs associated with surveying an area with this demonstrator's system. This includes data on equipment setup and calibration, site survey and any resurvey time, and downtime due to system malfunctions and maintenance requirements.

3.3.1 Survey Times

a. A government representative monitored and recorded all on-site activities, which were grouped into one of 11 categories. The first eight categories were chargeable to the system while the last three were not. Categorizing these activities provided insight into the technical and logistical aspects of the system. The times recorded in each category were then matched with the number of demonstrator personnel, assigned skill levels, and a consistent (across-vendor) salary to produce an estimate of the survey costs.

(1) Initial setup/mobilization. Started at the time when the demonstrator's equipment arrived at the survey site and stopped when the system was ready to acquire data.

(2) Daily setup/close-up. Monitored time spent mounting and dismounting the equipment each day.

(3) Instrument calibration. Recorded the amount of time used for daily quality assurance checks (e.g., sensors, GPS data, survey data quality).

(4) Data collection. Time spent surveying the test area.

(5) Downtime (nonsurvey time) for equipment/data checks. Covered time spent troubleshooting equipment or verifying survey tracks.

(6) Downtime (nonsurvey time) for equipment failure. Examples include replacing damaged cables, lost communication with base station, and any other failure that prevented surveying. Some weather-related failures fall into this category, for example, light-emitting diode (LED) displays darkened by the sun, wind creating waves too high to permit surveying, etc.

(7) Downtime (nonsurvey time) for maintenance. Battery replacement and memory downloads are typical examples.

(8) Demobilization. Commenced once the demonstrator completed the survey and concluded the final on-site check of the test data and ended when the equipment and personnel were ready to leave the site.

(9) Nonchargeable downtime for breaks and lunch. The demonstrator's company policy sets this standard.

(10) Nonchargeable downtime for weather-related causes (e.g., lightning, high wet-bulb heat index, and similar events).

(11) Nonchargeable downtime due to ATC range operating requirements. Danger zone conflicts, lack of support personnel, equipment, or other ATC-caused delays.

b. Appendix B contains the daily log sheets. Table 3-2 summarizes that information to provide insight into the operational, maintenance, and logistical aspects of the system.

TABLE 3-2. TIME ON-SITE

Date, 06	**20 Mar**	**21 Mar**	**22 Mar**	**23 Mar**	**24 Mar**	**Activity Totals, hr**
Activity (daily times recorded in minutes)						
Initial setup	445	-	-	-	-	**7.4**
Daily setup/close-up	40	150	110	75	40	**6.9**
Instrumentation calibration	-	25	25	-	30	**1.3**
Data collection	-	245	275	270	350	**19.0**
Equipment/data checks	-	-	-	-	85	**1.4**
Equipment failure	-	-	-	-	-	**0.0**
Maintenance	-	30	5	25	-	**1.0**
Demobilization	-	-	-	-	60	**1.0**
Breaks and lunch	-	15	-	10	20	**0.8**
Weather-related	-	-	155	-	-	**2.6**
ATC downtime	15	-	-	-	-	**0.3**
Daily total, hr	**8.3**	**7.8**	**9.5**	**6.3**	**9.8**	**41.7**

Note: Task times have been rounded to 5-minute increments.

3.3.2 On-Site Data Collection Costs

The times associated with the 11 activities have been reduced into the three basic components of the evaluation: initial setup, site survey, and pack-up (demobilization). Note that site survey time includes daily setup/stop time, collecting data, breaks/lunch, downtime due to equipment/data checks or maintenance, downtime due to failure, and downtime due to weather. This combines the actual survey cost with the demonstrator's associated on-site overhead costs.

A standardized estimate for labor costs associated with this effort was then calculated using the following job categories: supervisor ($95.00/hr), data analyst ($57.00/hr), and site support ($28.50/hr). The estimated costs are shown in Table 3-3.

TABLE 3-3. CALCULATED SURVEY COSTS

	No. of Persons	Hourly Wage	Hours	Cost
Initial Setup				
Supervisor	1	$95.00	7.4	$703.00
Data analyst	1	$57.00	7.4	$421.80
Site support	1	$28.50	7.4	$210.90
Subtotal				**$1335.70**
Site Survey				
Supervisor	1	$95.00	34.3	$3258.50
Data analyst	1	$57.00	34.3	$1955.10
Site support	1	$28.50	34.3	$977.55
Subtotal				**$6191.15**
Demobilization				
Supervisor	1	$95.00	1.0	$95.00
Data analyst	1	$57.00	1.0	$57.00
Site support	1	$28.50	1.0	$28.50
Subtotal				**$180.50**
Total on-site costs				**$7707.35**

3.4 COST ANALYSIS

The data collection process described above provides an on-site cost guide to compare the performance of this vendor with any other that has demonstrated at the shallow water site. It is not a true indicator of survey costs. Many other expenses have not been included, such as travel costs, per diem, off-site data processing and analysis, company overhead, and profit.

Calculating the area surveyed is done by plotting the raw GPS coordinates then combining the sensor swath (line spacing and associated overlap).

To determine the number of acres surveyed per day, the total number of hours spent at the test site (table 3-2) was divided by 8 (converts to 8-hour days). The number of acres was then divided by the number of 8-hour days. The cost per acre was determined by dividing the total survey costs (table 3-3) by the same number of acres. This information is summarized in Table 3-4.

TABLE 3-4. SURVEY COSTS

Area surveyed (acres[a])	4.25
Time on-site (8-hr days)	5.2
Calculated survey cost (U.S. dollars)	$7707
Acres per day	0.82
Cost per acre	$1813

[a]Acre = 4047 m^2.

Table 3.5 presents a comparison of Tetra Tech's survey costs with the EQT-ORD criteria.

TABLE 3-5. TEST RESULTS - CRITERIA COMPARISON

Metric	Threshold	Objective	CTC
Cost rate	$4000 per acre	$2000 per acre	$1813 per acre
Production rate	5 acres per day	50 acres per day	0.82 acres per day

SECTION 4. TECHNICAL PERFORMANCE RESULTS

4.1 AREA SURVEYED

4.1.1 Calculated Area

a. Both the test and scoring methodologies required the demonstrator to survey 100 percent of each of the four test areas (blind grid, open water, littoral, and deeper water). Scoring a partially surveyed area alters the ordnance and clutter sample sizes and test area boundaries and decreases the statistical confidence in the performance statements made for that area. Allowing partial scoring decreases the validity of performance comparisons made between multiple test areas for a single demonstrator and comparisons made between multiple demonstrators for a single test area.

b. Realizing that some systems may not be able to survey 100 percent of a given test area, a ranking system was established. The percent coverage for a given test area is determined by first plotting the raw GPS coordinates combined with the sensor swath (line spacing and associated overlap), calculating the area surveyed, and then comparing that surveyed area with the total test area.

$$\frac{\text{Section Surveyed}}{\text{Test Area Size}} \times 100 = \%\ \text{Surveyed}$$

c. The demonstrator's system is always scored against the complete ground truth for a given test area regardless of the percentage covered.

4.1.2 Area Assessment

The ranking system and survey results are presented in Table 4-1.

TABLE 4-1. SURVEY RANKING SYSTEM AND RESULTS

Ranking System		Survey Results, M882		
% Area Covered	**Ranking**	**Test Area**	**% Area Covered**	**Data Use**
95 to 100	Met	Blind grid	100	Direct comparison between systems and areas.
90 to 94	Generally met	Deeper water	94	Comparison between systems and areas. A small negative bias is contained in the reported numbers (bias not quantified in this report).
50 to 89	Partially met	Open water	84	Reported, not compared between systems or areas. A large negative bias is contained in the reported numbers (bias not quantified in this report).
		Littoral	74	
0 to 49	Not met			Not scored/not reported.

4.2 SYSTEM SCORING PROCEDURES

a. The scoring entities used in this program were predicated on knowing the composition and location of every detectable item in an area. The deeper water area is the one exception. Ground truth targets were placed in this area without a pre-survey and clearing operation. Therefore, only the system's probability of detection (P_d) was evaluated in this area.

b. The best indicator of survey performance is the blind grid. This area provides a statically valid, controlled environment in which the demonstrator must provide a response (ordnance, clutter, or blank) at each of the 644 locations. Comparison of the response and discrimination lists to the ground truth in this area both determines the range of ordnance the system can reliably detect and establishes the baseline to which system performance in all other test areas is measured.

c. The scoring terms and definitions, along with an explanation of the receiver operating characteristic (ROC) curve development and the chi-square analysis used in this report, are provided in Appendix C.

d. Demonstrator performance was scored in two stages: response and discrimination.

e. Response stage scoring evaluated the ability of the demonstrator's system to detect emplaced ground truth targets without regard to discriminating ordnance from clutter. In this stage, the GPS locations and signal strengths of all anomalies that the demonstrator deemed sufficient for further investigation and/or processing were reported. This list was generated with minimal processing, i.e., associating signal strength with GPS location, and included only signals that were above the system noise level.

f. The discrimination stage evaluated the demonstrator's ability to segregate ordnance from clutter. The same GPS locations reported in the response stage anomaly list were evaluated on the basis of the demonstrator's discrimination process (section 2.6). A discrimination stage list was generated and prioritized based on the demonstrator's determination that an anomaly was more likely to be ordnance rather than clutter. Typically, higher output values indicate a higher confidence that an ordnance item is present at a specified location. The demonstrator then specifies the threshold value for the prioritized ranking that provides optimum system performance. This value is the discrimination stage threshold.

g. Both the response and discrimination lists contain an identical number of potential target locations. They differ only in the priority ranking of the declarations.

h. Within both of these stages, the following entities were measured:

(1) P_d.

(2) Probability of false positive (P_{fp}).

(3) Probability of background alarm (P_{ba})/background alarm rate (BAR).

4.2.1 Deviations From Scoring Procedures

Foerster was responsible for the magnetometer data reduction and analysis. They use evaluation software called DATALINE, which provides a quality factor (0 to 100) to characterize the performance of the dipole fit routine for each object calculation. The quality factor is associated with a volume/diameter calculation and a visual evaluation of the magnetic anomaly map. Using both numerical values produced by the software and a visual interpretation of the dipole on the anomaly map, the analyst determines whether an object is scrap or an item of interest. If an item doses not exist at a given location, a quality factor number cannot be produced. This is only an issue for scoring in the blind grid area.

The minimally processed signal list and final dig list submitted by CTC/Foerster were both in accordance with the contract requirements. However, it was necessary for ATC to modify the blind grid dig list to fit the automated scoring routine. The first modification ATC made to the dig list was to include a zero value for all cell center locations that did not have an associated signal strength (quality factor number). This addressed the issue of not having a value at cell centers that were called "blank." The signal strengths and associated item calls for all other cell centers remained unchanged. Applying the standardized scoring rules produced the results shown in Table 4-2.

Calculated values assume that the number of detections is a binomially distributed random variable. Reported results are at the 90 percent reliability/95 percent confidence levels unless otherwise noted.

TABLE 4-2. STANDARDIZED SCORING (ZERO-FILLED) DETECTION SUMMARY

Metric	Overall	By Projectile Caliber				
		40 mm	**60 mm**	**81 mm**	**105 mm**	**155 mm**
Blind grid						
Response stage						
P_d	26.2%	31.0%	24.1%	20.7%	27.6%	27.6%
P_d lower 90% confidence	21.5%	19.7%	14.0%	11.2%	16.8%	16.8%
P_{fp}	31.0%					
P_{fp} lower 90% confidence	26.4%					
P_{ba}	28.6%					
Discrimination stage						
P_d	15.2%	24.1%	3.4%	0.0%	20.7%	27.6%
P_d lower 90% confidence	11.4%	14.0%	0.4%	0.0%	11.2%	16.8%
P_{fp}	8.6%					
P_{fp} lower 90% confidence	6.0%					
P_{ba}	0.6%					
Response noise level	4					
Discrimination threshold	4					

The P_d, P_{fp}, and P_{ba} values in the response stage of Table 4-2 are all within a few percentage points of each other. The same is seen for the P_d values across projectile calibers. Together, these findings indicate that the response of this instrument in detecting ferrous objects was no better than chance. Discrimination results at this point are meaningless.

ATC decided to reanalyze this system by moving away from the "signal strength"-based analysis of the results to the "signal-interpreted" results provided by this demonstrator. Along with a signal strength at each cell center that contained an item, the demonstrator also provided an interpretation of that signal, i.e., ordnance, clutter, or blank (no value). ATC had already assigned a value of 0 for blank locations and subsequently assigned a value of 1 for items Foerster identified as clutter and a 2 for items called ordnance. The response threshold was set at 0.5 and the discrimination threshold at 1.5. Rescoring this system with these values produced the results in Table 4-3.

TABLE 4-3. MODIFIED SCORING (ZERO-FILLED) DETECTION SUMMARY

Metric	Overall	By Projectile Caliber				
		40 mm	60 mm	81 mm	105 mm	155 mm
Blind grid						
Response stage						
P_d	56.6%	65.5%	6.9%	27.6%	82.8%	100.0%
P_d lower 90% confidence	50.9%	51.9%	1.8%	16.8%	70.3%	92.4%
P_{fp}	28.2%					
P_{fp} lower 90% confidence	23.7%					
P_{ba}	4.0%					
Discrimination stage						
P_d	55.2%	65.5%	6.9%	27.6%	75.9%	100.0%
P_d lower 90% confidence	49.5%	51.9%	1.8%	16.8%	62.8%	92.4%
P_{fp}	24.7%					
P_{fp} lower 90% confidence	20.5%					
P_{ba}	4.0%					
Response noise level	0.5					
Discrimination threshold	1.5					

The relationships between the P_d, P_{fp}, and P_{ba} values shown in the response stage in this table are indicative of a functional detection system. As would be expected, the P_d values also increased with projectile size in the 60- to 155-mm caliber range. An explanation for the high probability of detection for both 155- and 40-mm projectiles was provided in an email from Foerster (ref 3) ". . . Under the assumption of an 'average permeability' for ferrous ammunition, the magnetic moments are converted into a volume/diameter indication of a spherical shaped object of this specific permeability. This value can be used for size classification, after a calibration trial is performed.

"The following volume classification could be defined by means of the calibration lanes:

155 mm	12 . . . 20 liters
105 mm	1 . . . 5 liters
81 mm	1 . . . 6 liters
60 mm	1 . . . 3 liters
40 mm	< 0.2 liters . . . ."

The better-defined volumes for the smallest and largest ordnance items contributed to the higher probability of detection and classification for these extremes, whereas the overlapping volumes for the intermediate calibers contributed to the reduced detection and classification results. As shown later in this report, this trend holds true for the open water and littoral test areas as well.

Foerster did not identify (discriminate) cell contents by projectile caliber. The discrimination results in Tables 4-2 and 4-3 represent the percentage of each projectile population that was first recognized above the response stage noise threshold and then retained as being above the discrimination threshold. The relationship between the P_d, P_{fp}, and P_{ba} values shown in the discrimination stage are also indicative of a functional discrimination process.

The multiple signal processing and human interpretation steps that Foerster uses in the analysis and reporting of anomalies make such an analysis incompatible with the signal strength-based analytical procedure that is typically used to evaluate shallow water MEC detection systems. In the interest of accurately evaluating the performance of this system, ATC used the signal-interpreted values to measure this system's performance in the three other test areas as well; that is, regardless of signal strength, if an object was called "ordnance" in either the response or the discrimination stage, it remained in that category throughout the scoring process. All other standardized scoring rules applied.

4.2.2 ROC curves

Based on the entire range of ground truth targets used at this site, ROC curves were generated for both the response and discrimination stages. In both stages, the probability of detection versus false alarm rates was plotted. False alarms were divided into two groups: (1) anomalies corresponding to emplaced clutter items, thereby measuring the P_{fp}, and (2) anomalies not corresponding to any known item, termed background alarms (P_{ba}) in the blind grid area and BAR in all other areas.

The ROC curves for the response and discrimination stages for all areas surveyed are shown in Figures 6 through 13. Horizontal lines illustrate the system performance at the demonstrator's recommended noise level during the response stage, or discrimination threshold level in the discrimination stage. The point where the curve crosses the horizontal line defines the subset of targets that the demonstrator recommends digging.

Blind grid ROC curves showing both the signal strength and signal-interpreted results are shown in Figures 6 through 9. The slopes of the signal strength response curves in Figures 6 and 7 imply that the instrument responds as well to clutter and background alarms as it does to

ordnance. When the slopes of the discrimination curve in the same graphs are compared with those of the response curves, the improvement, based on the discrimination process, is readily apparent. The best performance of this system is reflected at the top end of the discrimination curve; however, the reported efficiency and rejection values are based on the demonstrator-provided signal-noise and discrimination thresholds. These values intersect the curve at a much lower point.

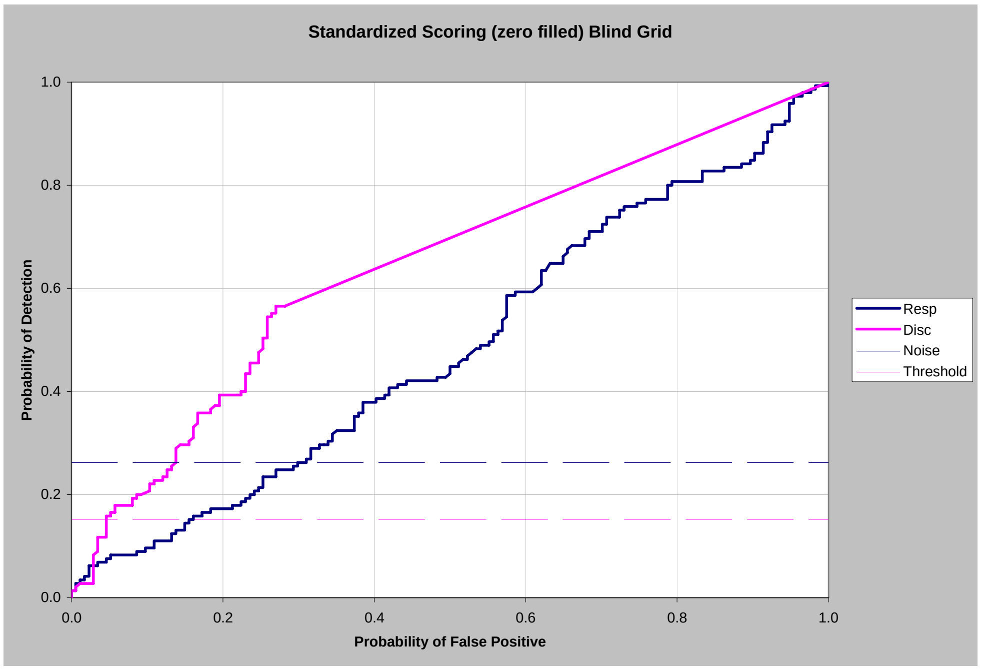

Figure 6. Standardized scoring - blind grid P_d versus P_{fp}.

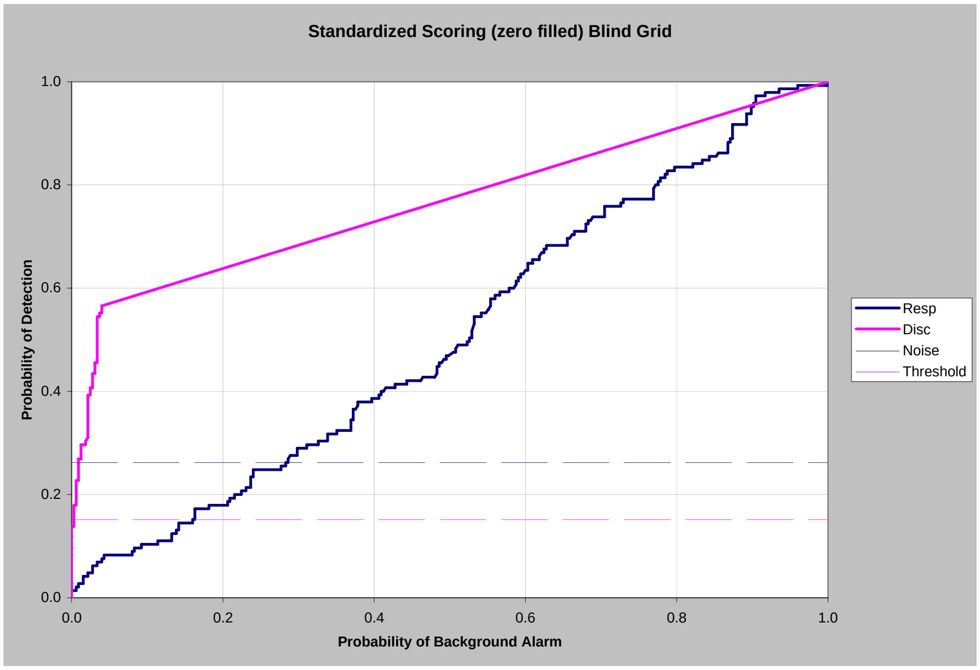

Figure 7. Standardized scoring - blind grid P_d versus P_{ba}.

The signal-interpreted ROC curves for the blind grid are shown in Figures 8 and 9. The curves shown in these figures are typical of those produced by a "mag-and-flag" operation. For the most part, the response and discrimination curves overlap each other. There is a small difference between the signal-noise and discrimination thresholds due to a few item classification changes going from the response stage to the discrimination stage. Two observations can be made when the signal strength and signal-interpreted sets of ROC curves are compared. The first is that the slope of the discrimination curves is essentially the same (table 4-4).

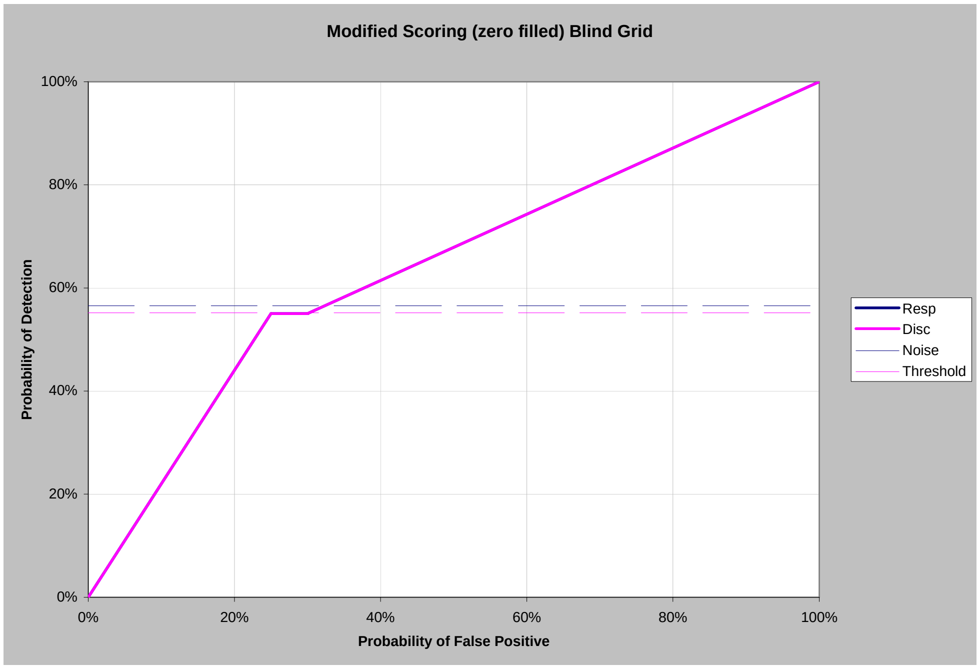

Figure 8. Modified scoring - blind grid P_d versus P_{fp}.

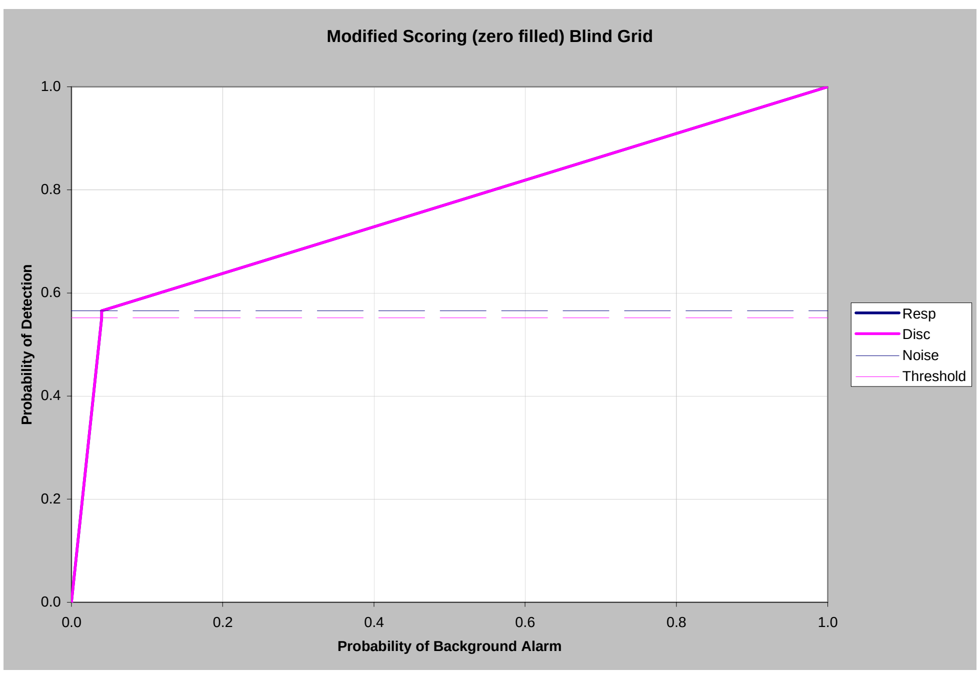

Figure 9. Modified scoring - blind grid P_d versus P_{ba}.

TABLE 4-4.   LEAST-SQUARED DISCRIMINATION SLOPE ANALYSIS

	Signal-Interpreted	Signal Strength
P_{fp}	$y = 2.2326x$ $R^2 = 1$	$y = 1.8402x + 0.0258$ $R^2 = 0.9797$
P_{ba}	$y = 13.793x$ $R^2 = 1$	$y = 11.902x + 0.1009$ $R^2 = 0.9461$

The second observation is that the signal-noise and discrimination thresholds are now located closer to the system's peak probability of detection and discrimination values. System efficiency measures the amount of detected ordnance retained by the discrimination process at a threshold of interest (i.e., the demonstrator's discrimination threshold). As the quantity of ordnance items that fall below this threshold increases, so does the efficiency rating of the system.

The ROC curves shown for the open water and littoral areas are based on the modified scoring results. Because the values provided by Foerster are identical in the response and discrimination stages, the noise and discrimination thresholds and the response and discrimination curves overlap each other in these graphs. These curves represent the best performance possible from the CTC/Foerster system.

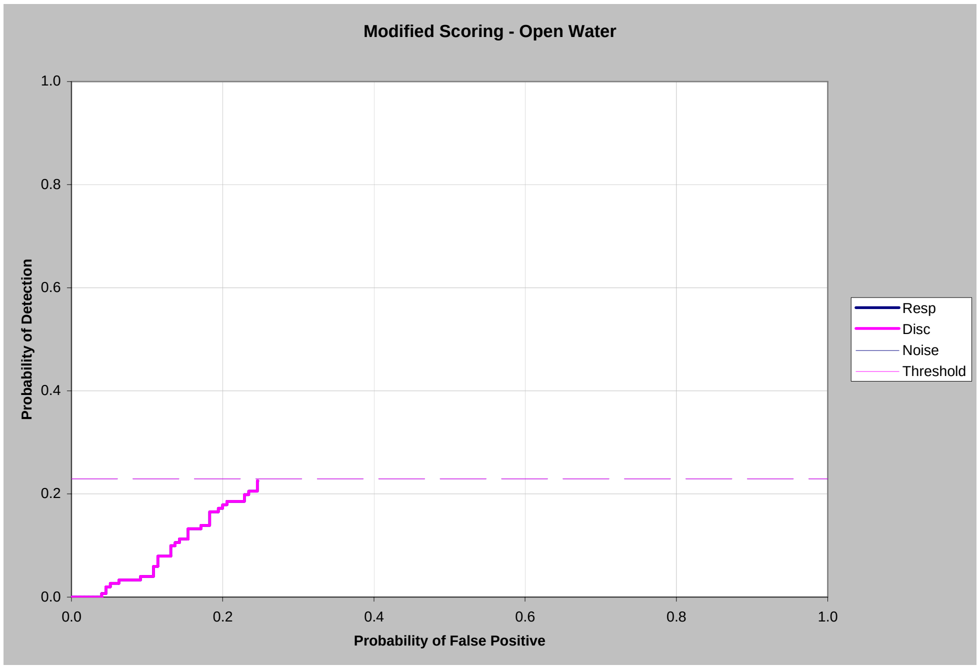

Figure 10.   Modified scoring - open water P_d versus P_{fp}.

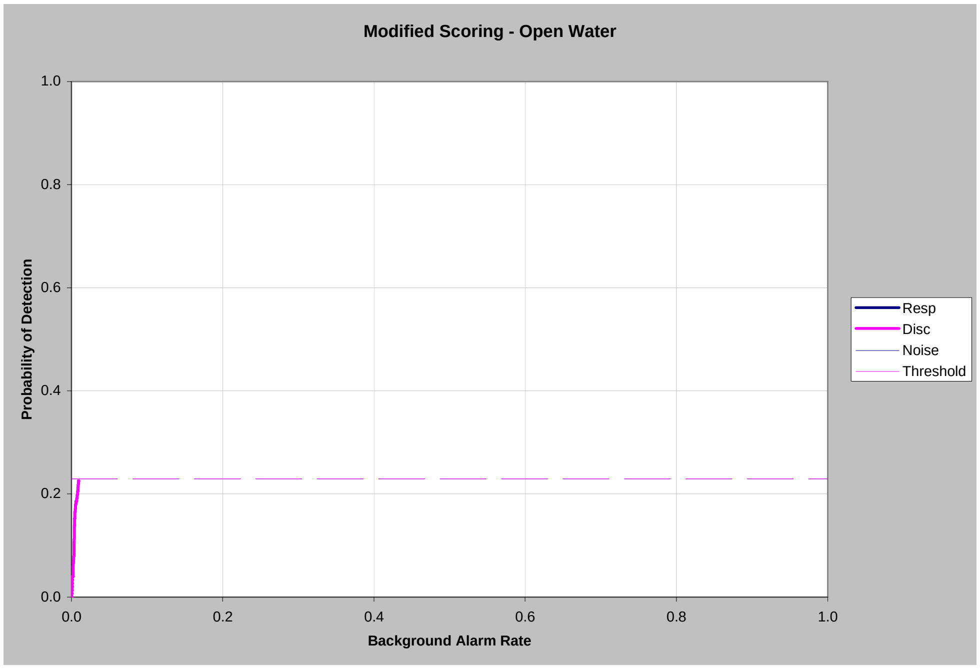

Figure 11. Modified scoring - open water P_d versus BAR.

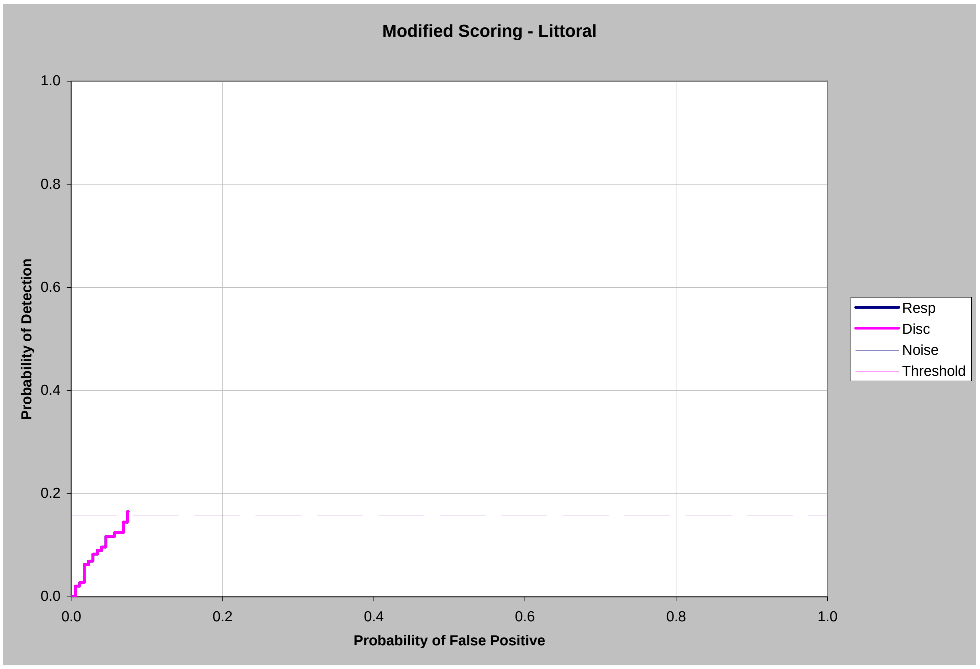

Figure 12. Modified scoring - littoral P_d versus $P_{fp.}$

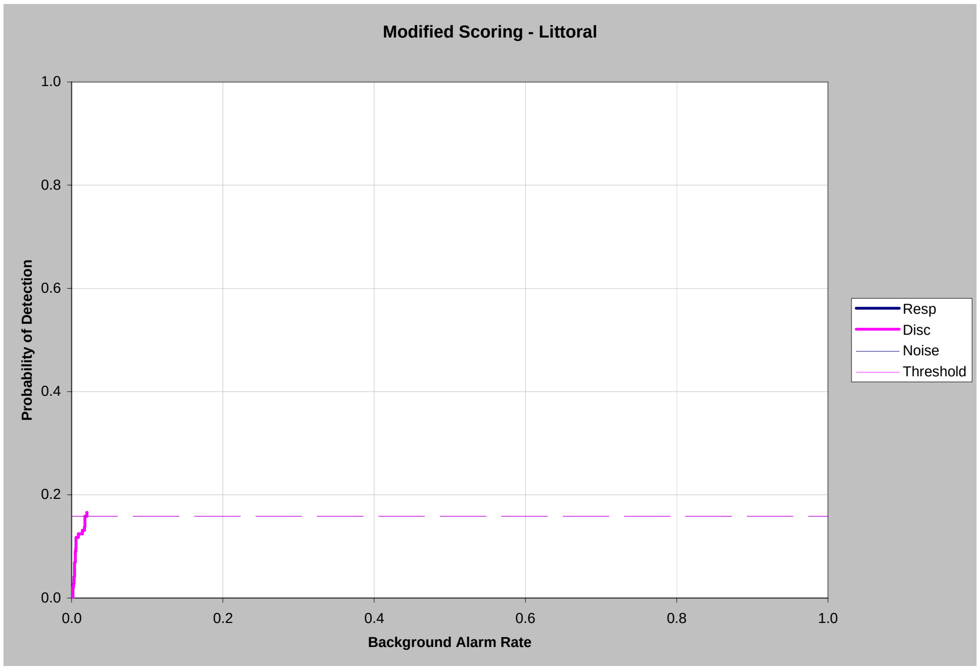

Figure 13. Modified scoring - littoral P_d versus BAR.

4.2.3 Detection Results

Detection results, broken out by stage, area surveyed, and ordnance size, are presented in Table 4-5. (The blind grid results are in tables 4-2 and 4-3) The results by size indicate how well the demonstrator did at detecting/discriminating ordnance of a given caliber. Overall results summarize ordnance detection over a given area. Calculated values assume that the number of detections is a binomially distributed random variable. Reported results are at the 90 percent reliability/95 percent confidence levels unless otherwise noted.

TABLE 4-5. MODIFIED SCORING SYSTEM DETECTION SUMMARY

Metric	Overall	By Projectile Caliber					
		40 mm	60 mm	81 mm	105 mm	155 mm	8 in.
Open water							
Response stage							
P_d	22.9%	27.6%	27.6%	20.7%	17.2%	20.0%	33.3%
P_d lower 90% confidence	18.6%	16.8%	16.8%	11.2%	8.6%	11.5%	9.3%
P_{fp}	20.2%						
P_{fp} lower 90% confidence	16.6%						
BAR m^{-2}	0.009						
Discrimination stage							
P_d	22.3%	24.1%	27.6%	20.7%	17.2%	20.0%	33.3%
P_d lower 90% confidence	18.0%	14.0%	16.8%	11.2%	8.6%	11.5%	9.3%
P_{fp}	19.7%						
P_{fp} lower 90% confidence	16.1%						
BAR m^{-2}	0.009						
Littoral region							
Response stage							
P_d	15.9%	24.1%	0.0%	6.9%	3.4%	44.8%	
P_d lower 90% confidence	12.0%	14.0%	0.0%	1.8%	0.4%	31.9%	
P_{fp}	6.9%						
P_{fp} lower 90% confidence	4.5%						
BAR m^{-2}	0.019						
Discrimination stage							
P_d	14.5%	24.1%	0.0%	6.9%	3.4%	37.9%	
P_d lower 90% confidence	10.8%	14.0%	0.0%	1.8%	0.4%	25.7%	
P_{fp}	6.9%						
P_{fp} lower 90% confidence	4.5%						
BAR m^{-2}	0.018						
Deeper water							
Response stage							
P_d	55.2%					55.2%	
P_d lower 90% confidence	41.7%					41.7%	
Discrimination stage							
P_d	55.2%					55.2%	
P_d lower 90% confidence	41.7%					41.7%	
Response noise level	4						
Discrimination threshold	4						

4.2.4 System Discrimination

Using the demonstrator's recommended setting, the items that were detected and correctly classified as ordnance were further evaluated as to whether the demonstrator could correctly identify the ordnance type. The list of ground truth ordnance items was provided to the demonstrator before testing.

CTC/Foerster's "dig list" discriminated between ordnance and clutter but not between ordnance types. The latter was an optional requirement.

4.2.5 System Effectiveness

Efficiency and rejection rates were calculated to quantify the discrimination ability at two specific points of interest on the ROC curve: the point where no decrease in P_d occured (i.e., the efficiency is, by definition, equal to 1) and the operator-selected threshold. These values are presented in Table 4-6.

TABLE 4-6. SIGNAL-INTERPERTED SCORING EFFICIENCY AND REJECTION RATES

	Efficiency	False Positive Rejection Rate	Background Alarm Rejection Rate
Blind Grid			
At operating point	0.98	0.12	0.00
With no loss of P_d	1.00	0.12	0.00
At operating point	0.58	0.72	0.98
With no loss of P_d	1.00	0.72	0.98
Open Water			
At operating point	0.97	0.02	0.10
With no loss of P_d	1.00	0.02	0.10
Littoral Region			
At operating point	0.91	0.00	0.04
With no loss of P_d	1.00	0.00	0.04

Note: Shaded values are based on signal strength analysis.

4.2.6 Chi-Square Analysis

A chi-square 2 x 2 Contingency Test for comparison between ratios was used to compare performance across the blind grid and deeper water test areas with regard to P_d^{res} and P_d^{disc}. A one-sided chi-square significance test at the 0.05 significance level was used. The intent of the comparison was to determine whether the features introduced in each test site had a degrading effect on the performance of the sensor system. These results are shown in Table 4-7.

TABLE 4-7. CHI-SQUARE SIGNIFICANCE TEST RESULTS

Metric	Overall	By Projectile Caliber				
		40 mm	60 mm	81 mm	105 mm	155 mm
Blind Grid - Deeper Water Comparison						
P_d^{res}	SIG					SIG
P_d^{disc}						

SIG = significant

4.2.7 Location Accuracy

The data points in the scatter graphs shown in Figures 14 and 15 represent the coordinates of ordnance items in the open water and littoral test areas that were first detected in the response stage within a 0.5-meter radius of their true positions and then correctly identified as ordnance in the discrimination stage. The maximum error represents the 0.5-meter detection limit. The mean error represents the statistical mean of the sample considered.

A visual assessment of the graphs indicates that the location error is a randomly distributed as opposed to a systematic error.

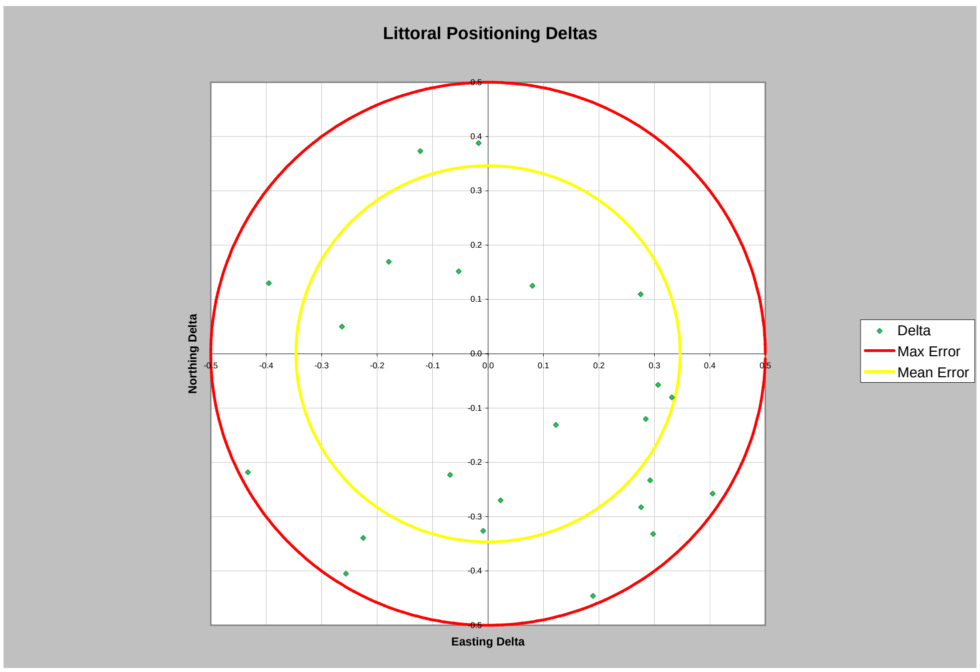

Figure 14. CTC/Foerster littoral positioning deltas.

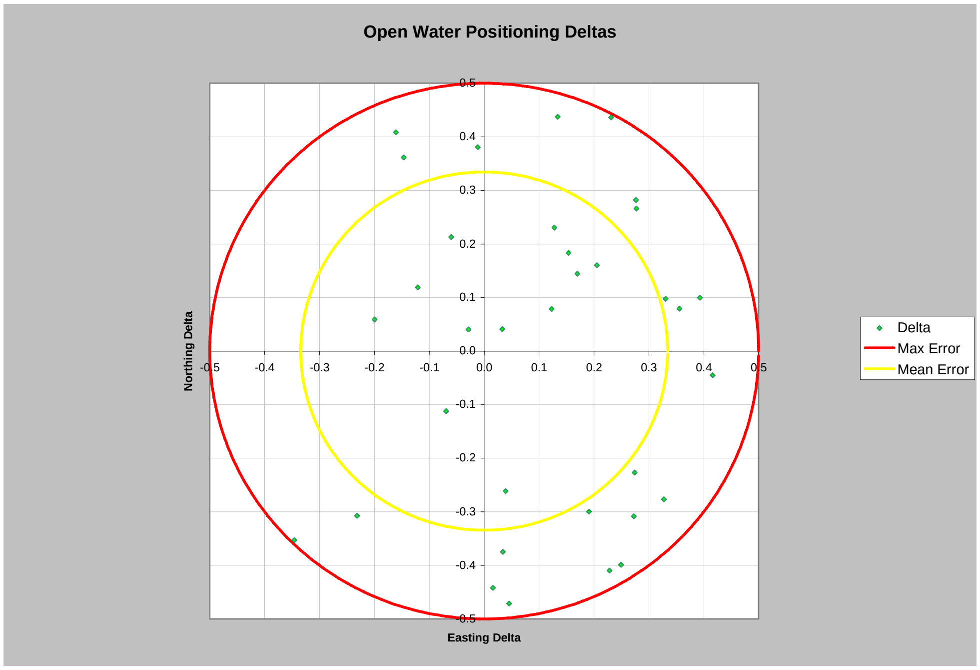

Figure 15. CTC/Foerster open water positioning deltas.

The comparison between the test results and the EQT-ORD criteria is presented in Table 4-8.

TABLE 4-8. TEST RESULTS - CRITERIA COMPARISON

Metric	Threshold	Objective	CTC by Area	
Detection	80% ordnance items buried to 1 foot and under 8 feet (2.4 m) of water.	95% ordnance items buried to 4 feet and under 8 feet (2.4 m) of water.	Blind grid	56.6%
			Open water	22.9%
			Littoral	15.9%
Discrimination	Rejection rate of 50% of emplaced non-UXO clutter.	Rejection rate of 90% of emplaced non-UXO clutter.	Blind grid	12%
			Open water	2%
			Littoral	0%
	Maximum false negative rate of 10%.	Maximum false negative rate of 0.5%.	Not assessed. An analytical procedure is not available to address this criterion.	
Reacquisition	Reacquire within 1 meter.	Reacquire within 0.5 meter.	The reported detection values are based on ordnance items identified within 0.5 meter of the geophysically referenced ground truth targets.	

Note: The blind grid and open water areas are in general accordance with the threshold requirements.

SECTION 5. APPENDIXES

APPENDIX A. TEST CONDITIONS LOG

ATMOSPHERIC CONDITIONS

Date, 06	Time, EDT	Average Wind Direction, deg	Average Wind Speed, km/h	Wind Direction Average Standard Deviation, deg	Peak Wind Speed, km/h	Average Temperature, °C
20 Mar	0700	1	2.3	22	5.1	-1.6
	0800	334	3.2	19	11.7	1.3
	0900	338	12.1	16	28.2	3.3
	1000	330	18.3	16	30.9	4.5
	1100	342	19.6	14	35.1	4.4
	1200	342	17.2	18	30.9	4.8
	1300	329	13.7	25	25.3	5.7
	1400	316	13.4	21	27.4	6.6
	1500	315	15.1	17	29.0	7.3
	1600	316	13.2	21	24.5	7.6
	1700	319	14.3	14	24.9	7.4
21 Mar	0700	30	5.8	13	11.3	-3.3
	0800	3	9.5	21	18.2	-2.3
	0900	11	11.4	21	20.8	-1.1
	1000	339	8.7	35	17.9	0.1
	1100	358	9.7	25	18.2	0.7
	1200	345	8.0	36	17.9	2.0
	1300	325	9.0	26	18.7	2.6
	1400	313	7.2	27	15.1	3.1
	1500	317	7.1	20	20.0	3.5
	1600	325	6.8	28	14.8	3.7
	1700	322	6.4	18	12.4	3.7
22 Mar	0700	327	16.7	12	31.9	0.2
	0800	331	23.2	13	43.5	0.6
	0900	331	27.7	14	45.9	1.6
	1000	333	29.5	13	48.3	2.9
	1100	331	25.7	16	42.2	4.2
	1200	319	22.4	16	43.6	5.2
	1300	316	23.5	17	39.6	5.0
	1400	308	20.1	18	36.5	6.0
	1500	307	19.0	18	34.4	6.6
	1600	315	19.3	14	35.7	6.5
	1700	320	19.6	19	33.0	6.3

Date, 06	Time, EDT	Average Wind Direction, deg	Average Wind Speed, km/h	Wind Direction Average Standard Deviation, deg	Peak Wind Speed, km/h	Average Temperature, °C
23 Mar	0700	278	12.9	11	21.6	2.2
	0800	298	13.5	14	24.5	3.3
	0900	327	13.0	18	26.4	4.7
	1000	332	16.6	12	31.4	5.6
	1100	336	15.1	16	24.9	6.1
	1200	313	13.0	21	25.4	6.9
	1300	309	10.9	27	22.0	8.1
	1400	288	12.2	22	24.8	8.8
	1500	297	13.0	20	24.6	9.2
	1600	311	13.8	20	25.7	9.6
	1700	321	14.0	13	25.4	9.2
24 Mar	0700	327	16.7	12	31.9	0.2
	0800	331	23.2	13	43.5	0.6
	0900	331	27.7	14	45.9	1.6
	1000	333	29.5	13	48.3	2.9
	1100	342	19.6	14	35.1	4.4
	1200	342	17.2	18	30.9	4.8
	1300	329	13.7	25	25.3	5.7
	1400	316	13.4	21	27.4	6.6
	1500	315	15.1	17	29.0	7.3
	1600	316	13.2	21	24.5	7.6
	1700	319	14.3	14	24.9	7.4

Note: The water conditions during the CTC survey were lost because of a malfunction in the portable tide gauge system. The water depth was measured against an elevation marker attached to the pier.

Company: CTC/Forester Date: 20 March 2006			Personnel: Josh Bowers, Tom Himmler, Myles Capen	
Start	**Stop**	**Remarks**	**Activity**	**Chargeable, min**
0825	0840	Arrived at test site. Safety briefing/questions.	Downtime (ATC)	15
0840	0900	Walked around pond for familiarization.	Initial setup	20
0900	1545	Attached the wooden framework to the aluminum boat. Attached trolling motor. Programmed into positioning system. Four sensors placed in polyvinyl chloride pipes that were sealed at the bottom. There was 0.5 meter of separation between the pipes (sensors).	Initial setup	405
1545	1605	Navigation practice.	Initial setup	20
1605	1645	End of day cleanup.	Daily close-up	40

Company: CTC/Forester Date: 21 March 2006			Personnel: Josh Bowers, Tom Himmler, Myles Capen	
Start	**Stop**	**Remarks**	**Activity**	**Chargeable, min**
0800	0930	Arrived at test site; began setup. Probes set to 6 feet for the blind grid area.	Daily setup	90
0930	0935	Calibration.	Calibration	5
0935	1155	Surveyed, concentrating on the blind grid. Wind light, waves calm.	Data collection	140
1155	1210	Replaced trolling motor battery.	Maintenance	15
1210	1225	Break.	Nonchargeable downtime	15
1225	1410	Blind grid survey complete.	Data collection	105
1410	1430	Took depth measurements in other areas of the pond to determine the level at which to set the sensors.	Calibration	20
1430	1445	Switched battery.	Maintenance	15
1445	1500	Continued survey.	Data collection	15
1500	1540	End of day cleanup.	Daily close-up	40

Company: CTC/Forester Date: 22 March 2006			Personnel: Josh Bowers, Tom Himmler, Myles Capen	
Start	**Stop**	**Remarks**	**Activity**	**Chargeable, min**
0800	0920	Setup. Plan was to take depth readings and set out buoys in preparation for survey.	Daily setup	80
0920	0945	Depth readings. Strong winds, 4- to 6-inch waves. The wind made maneuvering difficult (tide gauge not functioning).	Calibration	25
1020	1255	Stopped survey, wind too strong for the electric motor (55-lb thrust) Left site to look for a gas motor. Unsuccessful in locating a gas motor.	Weather	155
1255	1420	Resumed survey using the electric motor.	Data collection	205
1420	1425	Replaced motor battery.	Maintenance	5
1425	1535	Survey.	Data collection	70
1535	1605	End of day cleanup.	Daily close-up	30

Company: CTC/Forester Date: 23 March 2006			Personnel: Josh Bowers, Tom Himmler, Myles Capen	
Start	**Stop**	**Remarks**	**Activity**	**Chargeable, min**
0800	0850	Setup.	Daily setup	50
0850	1035	Survey.	Data collection	105
1035	1050	Changed battery.	Maintenance	15
1240	1250	Changed battery.	Maintenance	10
1250	1300	Lunch.	Nonchargeable downtime	10
1300	1545	Survey.	Data collection	165
1545	1610	End of day cleanup.	Daily close-up	25

Company: CTC/Foerster **Date: 24 March 2006**			**Personnel: Josh Bowers, Tom Himmler, Myles Capen**	
Start	**Stop**	**Remarks**	**Activity**	**Chargeable, min**
0800	0840	Setup. Plan was to survey the deeper water and littoral zones.	Daily setup	40
0840	1010	Littoral survey complete.	Data collection	90
1010	1035	Lowered probe depth to 9 feet for deeper water area.	Downtime	25
1035	1205	Survey.	Data collection	90
1205	1220	Changed motor battery.	Maintenance	15
1220	1330	Survey.	Data collection	70
1330	1355	Reset probes to 2 feet.	Calibration	25
1355	1415	Lunch.	Nonchargeable downtime	20
1415	1555	Surveyed the littoral zone.	Data collection	100
1555	1630	Repositioned probes to survey the calibration lanes.	Calibration	35
1630	1700	Surveyed calibration lanes.	Calibration	30
1700	1800	Demobilization.	Demobilization	60

APPENDIX C. TERMS AND DEFINITIONS

GENERAL DEFINITIONS

Anomaly: Location of a system response deemed to warrant further investigation by the demonstrator for consideration as an emplaced ordnance item.

Detection: An anomaly location that is within R_{halo} of an emplaced ordnance item.

Munitions and Explosives of Concern (MEC): Specific categories of military munitions that may pose unique explosive safety risks, including UXO as defined in 10 USC 101(e)(5), DMM as defined in 10 USC 2710(e)(2) and/or munitions constituents (e.g., TNT, RDX) as defined in 10 USC 2710(e)(3) that are present in high enough concentrations to pose an explosive hazard.

Emplaced Ordnance: An ordnance item buried by the government at a specified location in the test site.

Emplaced Clutter: A clutter item (i.e., nonordnance item) buried by the government at a specified location in the test site.

R_{halo}: A predetermined radius about the periphery of an emplaced item (clutter or ordnance) within which a location identified by the demonstrator as being of interest is considered to be a response from that item. For the purpose of this program, a circular halo 0.5 meters in radius will be placed around the center of the object for all clutter and ordnance items less than 0.6 meters in length. When ordnance items are longer than 0.6 meters, the halo becomes an ellipse where the minor axis remains 1 meter and the major axis is equal to the projected length of the ordnance onto the ground plane plus 1 meter.

Response Stage Noise Level: The level that represents the point below which anomalies are not considered detectable. Demonstrators are required to provide the recommended noise level for the blind grid test area.

Discrimination Stage Threshold: The demonstrators select the threshold level that they believe provides optimum performance of the system by retaining all detectable ordnance and rejecting the maximum amount of clutter. This level defines the subset of anomalies the demonstrator would recommend digging based on discrimination.

Binomially Distributed Random Variable: A random variable of the type that has only two possible outcomes, say, success and failure, and is repeated for n independent trials, with the probability p of success and the probability 1-p of failure being the same for each trial. The number of successes x observed in the n trials is an estimate of p and is considered to be a binomially distributed random variable.

RESPONSE STAGE DEFINITIONS

Response Stage Probability of Detection (P_d^{res}): P_d^{res} = (No. of response stage detections)/(No. of emplaced ordnance in the test site).

Response Stage False Positive (fp^{res}): An anomaly location that is within R_{halo} of an emplaced clutter item.

Response Stage Probability of False Positive (P_{fp}^{res}): P_{fp}^{res} = (No. of response stage false positives)/(No. of emplaced clutter items).

Response Stage Background Alarm: An anomaly in a blind grid cell that contains neither emplaced ordnance nor an emplaced clutter item. An anomaly location in the open water or littoral scenarios that is outside R_{halo} of any emplaced ordnance or emplaced clutter item.

Response Stage Probability of Background Alarm (P_{ba}^{res}): blind grid only: P_{ba}^{res} = (No. of response stage background alarms)/(No. of empty grid locations).

Response Stage Background Alarm Rate (BAR^{res}): open water only: BAR^{res} = (No. of response stage background alarms)/(arbitrary constant).

Note that the quantities P_d^{res}, P_{fp}^{res}, P_{ba}^{res}, and BAR^{res} are functions of t^{res}, the threshold applied to the response stage signal strength. These quantities can, therefore, be written as $P_d^{res}(t^{res})$, $P_{fp}^{res}(t^{res})$, $P_{ba}^{res}(t^{res})$, and $BAR^{res}(t^{res})$.

DISCRIMINATION STAGE DEFINITIONS

Discrimination: The application of a signal processing algorithm or human judgment to response stage data that discriminates ordnance from clutter. Discrimination should identify anomalies that the demonstrator has high confidence correspond to ordnance, as well as those that the demonstrator has high confidence correspond to nonordnance or background returns. The former should be ranked with highest priority and the latter with lowest.

Discrimination Stage Probability of Detection (P_d^{disc}): P_d^{disc} = (No. of discrimination stage detections)/(No. of emplaced ordnance in the test site).

Discrimination Stage False Positive (fp^{disc}): An anomaly location that is within R_{halo} of an emplaced clutter item.

Discrimination Stage Probability of False Positive (P_{fp}^{disc}): P_{fp}^{disc} = (No. of discrimination stage false positives)/(No. of emplaced clutter items).

Discrimination Stage Background Alarm: An anomaly in a blind grid cell that contains neither emplaced ordnance nor an emplaced clutter item. An anomaly location in the open water or littoral scenarios that is outside R_{halo} of any emplaced ordnance or emplaced clutter item.

Discrimination Stage Probability of Background Alarm (P_{ba}^{disc}): P_{ba}^{disc} = (No. of discrimination stage background alarms)/(No. of empty grid locations).

Discrimination Stage Background Alarm Rate (BAR^{disc}): BAR^{disc} = (No. of discrimination stage background alarms)/(arbitrary constant).

Note that the quantities P_d^{disc}, P_{fp}^{disc}, P_{ba}^{disc}, and BAR^{disc} are functions of t^{disc}, the threshold applied to the discrimination stage signal strength. These quantities can, therefore, be written as $P_d^{disc}(t^{disc})$, $P_{fp}^{disc}(t^{disc})$, $P_{ba}^{disc}(t^{disc})$, and $BAR^{disc}(t^{disc})$.

RECEIVER OPERATING CHARACERISTIC (ROC) CURVES

ROC curves at both the response and discrimination stages can be constructed based on the above definitions. The ROC curves plot the relationship between P_d versus P_{fp} and P_d versus BAR or P_{ba} as the threshold applied to the signal strength is varied from its minimum (t_{min}) to its maximum (t_{max}) value.[1] Figure A-1 shows how P_d versus P_{fp} and P_d versus BAR are combined into ROC curves. Note that the "res" and "disc" superscripts have been suppressed from all the variables for clarity.

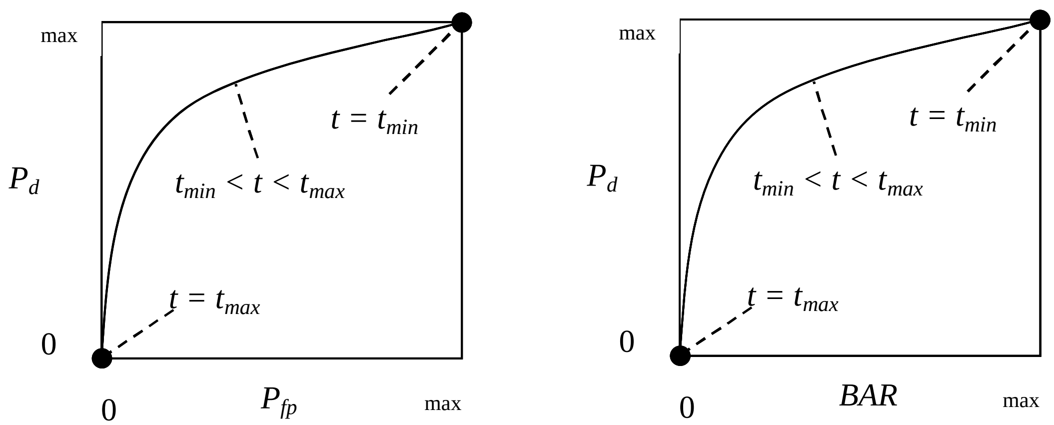

Figure A-1. ROC curves for open-site testing. Each curve applies to both the response and discrimination stages.

[1]Strictly speaking, ROC curves plot the P_d versus P_{ba} over a predetermined and fixed number of detection opportunities (some of the opportunities are located over ordnance and others are located over clutter or blank spots). In an open water scenario, each system suppresses its signal strength reports until some bare-minimum signal response is received by the system. Consequently, the open water ROC curves do not have information from low-signal output locations, and, furthermore, different contractors report their signals over a different set of locations on the ground. These ROC curves are thus not true to the strict definition of ROC curves as defined in textbooks on detection theory. Note, however, that the ROC curves obtained in the blind grid test sites are true ROC curves.

METRICS TO CHARACTERIZE THE DISCRIMINATION STAGE

The demonstrator is also scored on efficiency and rejection ratio, which measure the effectiveness of the discrimination stage processing. The goal of discrimination is to retain the greatest number of ordnance detections from the anomaly list while rejecting the maximum number of anomalies arising from nonordnance items. The efficiency measures the amount of detected ordnance retained by the discrimination, while the rejection ratio measures the fraction of false alarms rejected. Both measures are defined relative to the entire response list, i.e., the maximum ordnance detectable by the sensor and its accompanying false positive rate or background alarm rate.

Efficiency (E): $E = P_d^{disc}(t^{disc})/P_d^{res}(t_{min}^{res})$: measures (at a threshold of interest), the degree to which the maximum theoretical detection performance of the sensor system (as determined by the response stage t_{min}) is preserved after application of discrimination techniques. Efficiency is a number between 0 and 1. An efficiency of 1 implies that all of the ordnance initially detected in the response stage was retained at the specified threshold in the discrimination stage, t^{disc}.

False Positive Rejection Rate (R_{fp}): $R_{fp} = 1 - [P_{fp}^{disc}(t^{disc})/P_{fp}^{res}(t_{min}^{res})]$: measures (at a threshold of interest) the degree to which the sensor system's false positive performance is improved over the maximum false positive performance (as determined by the response stage t_{min}). The rejection rate is a number between 0 and 1. A rejection rate of 1 implies that all emplaced clutter initially detected in the response stage was correctly rejected at the specified threshold in the discrimination stage.

Background Alarm Rejection Rate (R_{ba}):

Blind grid: $R_{ba} = 1 - [P_{ba}^{disc}(t^{disc})/P_{ba}^{res}(t_{min}^{res})]$
Open water: $R_{ba} = 1 - [BAR^{disc}(t^{disc})/BAR^{res}(t_{min}^{res})])$

Measures the degree to which the discrimination stage correctly rejects background alarms initially detected in the response stage. The rejection rate is a number between 0 and 1. A rejection rate of 1 implies that all background alarms initially detected in the response stage were rejected at the specified threshold in the discrimination stage.

CHI-SQUARE COMPARISON EXPLANATION

The chi-square test for differences in probabilities (or 2 x 2 contingency table) is used to analyze two samples drawn from two different populations to see if both populations have the same or different proportions of elements in a certain category. More specifically, two random samples are drawn, one from each population, to test the null hypothesis that the probability of event A (some specified event) is the same for both populations (ref 4, pages 144 through 151).

A one-sided 2 x 2 contingency table is used in the Shallow Water Site Program to compare each area (open water, littoral, deep water) to the blind grid since each area introduces a water feature that makes it potentially more difficult to survey than the blind grid. The one-sided 2 x 2 contingency table is used to determine if there is reason to believe that the proportion of

ordnance correctly detected/discriminated by demonstrator X's system is significantly degraded by the more challenging feature introduced. A two-sided 2 x 2 contingency table is used to compare performance between any two of the test sites other than the blind grid, to determine if there is reason to believe that the proportion of ordnance correctly detected/discriminated by demonstrator X's system is significantly different between those two test sites.

The test statistic of the 2 x 2 contingency table is the chi-square distribution with one degree of freedom. For the one-sided test, a significance level of 0.05 is chosen, which sets a critical decision limit of 3.84 from the chi-square distribution with one degree of freedom. It is a critical decision limit because if the test statistic calculated from the data exceeds this value, the two proportions tested will be considered significantly different. If the test statistic calculated from the data is less than this value, the two proportions tested will be considered not significantly different.

An exception must be applied when either a 0 or 100 percent success rate occurs in the sample data. The chi-square test cannot be used in these instances. Instead, Fisher's Exact Test is used, and the critical decision limit is the chosen significance level, which is 0.05 for one-sided tests and 0.10 for two-sided tests. With Fisher's test, if the test statistic (p-value) is less than the critical value, then the null hypothesis of similar performance is rejected in favor of the alternative hypothesis: significantly greater than for the one-sided case or significantly different for the two-sided case.

Shallow water UXO Detection Test Site examples, where blind grid results are compared to those from the open water and littoral sites and the nongrid sites (open water and littoral) are compared to each other as follows. It should be noted that a significant result does not prove a cause and effect relationship exists between the change in survey area and sensor performance; however, it does serve as a tool to indicate that one data set reflects relatively degraded system performance of a large enough scale than can be accounted for merely by chance or random variation. Note also that a result that is not significant indicates that there is not enough evidence to declare that anything more than chance or random variation within the same population is at work between the two data sets being compared.

Demonstrator X achieves the following overall results after surveying each of the three areas using the same system (results indicate the number of ordnance detected divided by the number of ordnance emplaced):

	Blind Grid	Open Water	Littoral
P_d^{res}	100/100 = 1.0	8/10 = .80	20/33 = .61
P_d^{disc}	80/100 = 0.80	6/10 = .60	8/33 = .24

P_d^{res}: BLIND GRID versus OPEN WATER. Using the example data above to compare probabilities of detection in the response stage, all 100 ordnance out of 100 emplaced ordnance items were detected in the blind grid while 8 ordnance out of 10 emplaced were detected in the open water. Fisher's test must be used since a 100 percent success rate occurs in the data. Fisher's test uses the four input values to calculate a test statistic (p-value) of 0.0075 that is compared against the critical value of 0.05. Since the test statistic is less than the critical value,

the smaller response stage detection rate (0.80) is considered to be significantly less at the 0.05 level of significance. While a significant result does not prove a cause and effect relationship exists between the change in survey area and degradation in performance, it does indicate that the detection ability of demonstrator X's system seems to have been degraded in the open water relative to results from the blind grid using the same system.

P_d^{disc}: BLIND GRID versus OPEN WATER. Using the example data above to compare probabilities of detection in the discrimination stage, 80 out of 100 emplaced ordnance items were correctly discriminated as ordnance in blind grid testing while 6 out of 10 emplaced ordnance items were correctly discriminated as such in open water testing. Those four values are used in the chi-square Contingency Test to calculate a test statistic of 1.12. Since the test statistic is less than the critical value of 3.84, the two discrimination stage detection rates are considered to be not significantly different at the 0.05 level of significance.

P_d^{res}: BLIND GRID versus LITTORAL. Using the example data above to compare probabilities of detection in the response stage, 100 out of 100 and 20 out of 33 are used to calculate a test statistic (< 0.000) that is compared against the critical value of 0.05. Since the test statistic is less than the critical value, the smaller response stage detection rate (0.61) is considered to be significantly less at the 0.05 level of significance.

P_d^{disc}: BLIND GRID versus LITTORAL. Using the example data above to compare probabilities of detection in the discrimination stage, 80 out of 100 and 8 out of 33 emplaced ordnance items were correctly discriminated as such in open water testing. Those four values are used to calculate a test statistic of 32.01. Since the test statistic is greater than the critical value of 3.84, the smaller discrimination stage detection rate (0.24) is considered to be significantly less at the 0.05 level of significance.

P_d^{res}: OPEN WATER versus LITTORAL. Using the example data above to compare probabilities of detection in the response stage, 8 out of 10 and 20 out of 33 are used to calculate a test statistic of 0.56. Since the test statistic is less than the critical value of 2.71, the two response stage detection rates are considered to be not significantly different at the 0.10 level of significance.

P_d^{disc}: OPEN WATER versus LITTORAL. Using the example data above to compare probabilities of detection in the discrimination stage, 6 out of 10 and 8 out of 33 are used to calculate a test statistic of 2.98. Since the test statistic is greater than the critical value of 2.71, the two discrimination stage detection rates are considered to be significantly different at the 0.10 level of significance. While a significant result does not prove a cause and effect relationship exists between the change in survey area and change in performance, it does indicate that the ability of Demonstrator X to correctly discriminate seems to have been degraded by features of the littoral area relative to results from the open water using the same system.

APPENDIX D. REFERENCES

1. Environmental Quality Technology - Operational Requirements Document (EQT-ORD) for: A(1.6.a): UXO Screening, Detection and Discrimination.

2. Technical Management Plan, Detection and Discrimination Demonstration of a Fluxgate Vertical Gradient Magnetometer at the Aberdeen Shallow Water Test Site. Submitted in response to the BAA W91ZLK-04-R-0001, by Concurrent Technologies Corporation, 30 August 2005.

3. *Email*: 28 June 2006, sent from Mr. Thomas Himmler (Foerster GmbH & Co. KG) through Mr. Josh Bowers (bowersr@ctc.com) to Mr. Gary Rowe (gary.rowe@atc.army.mil) regarding SWDS Scoring.

4. Practical Nonparametric Statistics, W.J. Conover, John Wiley & Sons, 1980, pages 144 through 151.

APPENDIX E. ABBREVIATIONS

APG	=	Aberdeen Proving Ground
ASCII	=	American Standard Code for Information Interchange
ATC	=	U.S. Army Aberdeen Test Center
BAA	=	Broad Agency Announcement
BAR	=	background alarm rate
CTC	=	Concurrent Technologies Corporation
DGPS	=	Differential Global Positioning System
DMM	=	discarded military munitions
EQT	=	Army Environmental Quality Technology Program
EQT-ORD	=	Environmental Quality Technology - Operational Requirements Document
ERDC	=	U.S. Army Corps of Engineers Engineering, Research and Development Center
ESTCP	=	Environmental Security Technology Certification Program
GPS	=	Global Positioning System
LED	=	light-emitting diode
MEC	=	munitions and explosives of concern
METDC	=	Military Environmental Technology Demonstration Center
NMEA	=	National Marine Electronics Association
P_{ba}	=	probability of background alarm rate
P_d	=	probability of detection
P_d^{disc}	=	probability of detection, discrimination stage
P_d^{res}	=	probability of detection, response stage
P_{fp}	=	probability of false positive
P_{fp}^{disc}	=	probability of false positive, discrimination stage
P_{fp}^{res}	=	probability of false positive, response stage
POC	=	point of contact
QA	=	quality assurance
QC	=	quality control
ROC	=	receiver operating characteristic
SERDP	=	Strategic Environmental Research and Development Program
USAEC	=	U.S. Army Environmental Command
UXO	=	unexploded ordnance

APPENDIX F. DISTRIBUTION LIST

DTC Project No. 8-CO-160-UXO-016

Addressee	No. of Copies
Commander	
U.S. Army Aberdeen Test Center	
ATTN: CSTE-DTC-AT-SL-E (Mr. Gary Rowe)	1
(Library)	1
CSTE-DTC-AT-CS-R	1
Aberdeen Proving Ground, MD 21005-5059	
Booz Allen Hamilton	
ATTN: Mr. Patrick McDonnell	1
4692 Millennium Drive, Suite 200	
Belcamp, MD 21017-1535	
Concurrent Technologies Corporation (CTC)	
ATTN: Mr. Josh Bowers	1
100 CTC Drive	
Johnstown, PA 15904-1935	
Defense Technical Information Center	PDF
8725 John J. Kingman Road, Suite 0944	
Fort Belvoir, VA 22060-6218	

Secondary, "hard-copy" distribution is controlled by Commander, U.S. Army Environmental Command, ATTN: SFIM-AEC-ATT.